DOCUMENT RESUME

ED 470 590 TM 034 609

AUTHOR	Boston, Carol
TITLE	Effect Size and Meta-Analysis. ERIC Digest.
INSTITUTION	ERIC Clearinghouse on Assessment and Evaluation, College Park, MD.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	EDO-TM-02-08
PUB DATE	2002-12-00
NOTE	4p.
CONTRACT	ED-99-CO-0032
AVAILABLE FROM	ERIC CLearinghouse on Assessment and Evaluation, 1129 Shriver Laboratory, University of Maryland, College Park, MD 20742. Tel: 800-464-3742 (Toll Free); Web site: http://ericae.net.
PUB TYPE	ERIC Publications (071) -- ERIC Digests in Full Text (073)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Comparative Analysis; *Effect Size; *Meta Analysis; *Statistical Distributions
IDENTIFIERS	ERIC Digests; *Standard Deviation

ABSTRACT

This Digest provides a review and applications of the concepts of normal distribution, standard deviation, effect size, and translation of effect size into percentile gain, the foundations for the understanding of meta analytic results. Meta analysis is a statistical technique that enables the results of a number of studies to be combined to determine the average effect of a given technique. Comparisons can then be made about the relative effectiveness of various techniques for increasing student achievement. Meta analysis involves identifying, collecting, reviewing, coding, and interpreting scientific research studies. Study outcomes are translated to a common metric to allow results to be compared. Following the explanation of concepts important to meta analysis, the Digest provides an example and lists some resources for further information and tips for locating examples of meta analysis. (Contains 10 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made

ED 470 590

Effect Size and Meta-Analysis

Carol Boston
ERIC Clearinghouse on Assessment and Evaluation

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM034609

ERIC Digest EDO-TM-02-08 December 2002

BEST COPY AVAILABLE

Effect Size and Meta-Analysis

Carol Boston
ERIC Clearinghouse on Assessment and Evaluation

Given the growing demand for "evidenced-based research" to guide educational interventions, interest in the research technique of *meta-analysis* has surged. Developed by Gene Glass in the mid-1970s, meta-analysis is a statistical technique that enables the results from a number of studies to be combined to determine the average effect of a given technique. Comparisons can then be made about the relative effectiveness of various techniques for increasing student achievement.

Where traditional literature reviews often present a narrative, chronological look at a small subset of studies deemed by an author to be relevant to the question at hand, meta-analysis is a more exacting and objective process that involves identifying, collecting, reviewing, coding, and interpreting scientific research studies. Studies are typically coded according to such categories as publication in a peer-reviewed journal, sample size and composition, control group, use of randomization, research methodology, type of intervention, and length of intervention. Study outcomes are translated to a common metric, called an effect size, to allow results to be compared.

This Digest provides a review and applications of the concepts of normal distribution, standard deviation, effect size, and translation of effect size into percentile gain—foundations for the understanding of meta-analytic results. It also includes resources for further exploration.

Basics of Research Synthesis

Statistically, students' achievement scores tend to be distributed according to the well-known "bell curve," also known as *normal distribution*. In other words, the majority of scores are clustered around the mid-point of the scale, or distributed symmetrically around the mean, with fewer scores occurring as the distance from the mean increases according to a specific mathematical equation. *Standard deviation* is the measurement of how scores are clustered or dispersed in relation to the mean. It is a measure of variability, something akin to an average distance from the mean.

Normal distribution has a range of about three standard deviations above the mean and three standard deviations below the mean. In graphic terms, envision a bell-shaped curve divided in half at the highest part (the mean score), then add two more vertical lines at equal intervals on each side. About 68 percent of the population can be expected to lie within the first standard deviation on either side of the mean (34 percent on each side). About 95 percent of the population will lie within +/- two standard deviations, and 99 percent of the population will lie within +/- three standard deviations. To give an extremely simplified example, assume subject knowledge is normally distributed among 100 students, on a test of 100 items with a mean of 50 and a standard deviation of 20. About 34 students would be expected to score between 50 and 70, about 14 students would score between 71 and 90, and about 2 would score between 91 and 100. Fifty of the students would, of course, score below the mean, with 34 scoring between 30 and 50, 14 scoring between 10 and 29, and two students scoring between 0 and 9 points.

In order to show whether a particular technique or intervention helps raise student achievement on a test, a researcher would translate the results of a given study into a unit of measurement referred to as an *effect size*. An effect size expresses the increase or decrease in achievement of the experimental group (the group of students who are exposed to a specific instructional technique) in standard deviation units.

For example, suppose that the effect size computed for a specific study is 1.0. This means that the average score for students in the experimental group is 1.0 standard deviation higher than the average scores of students in the control group. In other words, a student at the 50th percentile in the experimental group would be one standard deviation higher than a student at the 50th percentile in the control group. A study that shows an effect size of 1.0 thus means a percentile gain of 34 points—one standard deviation above the mean encompasses 34 percent of the scores–provided one can assume the average for the group is the 50th percentile.

As Marzano, Pickering, and Pollock (2001) note, "Being able to translate effect sizes into percentile gains provides for a dramatic interpretation of the possible benefits of a given instructional strategy." By way of example, they report that Redfield and Rousseau (1981) analyzed 14 studies on the classroom use of higher level questions and computed the average effect size of those studies to be .73. This means that the average student who was exposed to higher level questioning strategies scored .73 standard deviations above the scores of the average student who was not exposed to higher level questioning strategies. Transforming effect sizes to percentile gains through statistical conversion shows that an effect size of .73 represents a percentile gain of about 27 points (p. 6).

Another way of interpreting effect size is to consider an effect size of .20 as small, an effect size of .50 as medium, and an effect size of .80 as large (Cohen, 1988). While these are accepted rules of thumb, the importance of an effect size magnitude is, in the end, a judgment call.

Identifying Effective Instructional Techniques Through Meta-Analyses

Meta-analysis is a widely accepted technique for summarizing studies and exploring relationships. The National Reading Panel convened by the National Institute of Child Health and Human Development and the U.S. Department of Education, for example, used the technique to assess the state of research-based knowledge about teaching children to read (NICHHD, 2000).

Researchers at the Mid-Continent Regional Educational Laboratory undertook a study to identify broad strategies with a high probability of enhancing student achievement for all students in all subject areas at all grade levels. A detailed look at these instructional strategies, their basis in research, and their classroom application, is provided in the handbook, *Classroom Instruction That Works: Research-Based Strategies for Increasing Student Achievement* (Marzano, Pickering, and Pollock, 2001). In the area of homework and practice, for example, the average effect size was .77—for a percentile gain of 28--based on 134 studies. The reported standard deviation of .36 tells how different those 134 studies were. One meta-analysis of homework involving research up to 1988 reported effect sizes of homework to be .15 for grades 4 to 6, .31 for grades 7 to 9, and .64 for grades 10 to 12 (Cooper, 1989). According to this study, whereas homework in high school produces a gain of about 24 percentile points, homework in the middle grades produces a gain of only 12 percentile points, and homework in grades 4 to 6 has a relatively small effect–a percentile gain of 6 points on student achievement. More recent studies have shown beneficial results for elementary school students as young as second grade (see, for example, Cooper, Lindsay, Nye, and Greathouse, 1998; Cooper, Valentine, Nye, and Lindsay, 1999).

Walberg (1999) found that the effects of homework vary greatly, depending on the feedback a teacher provides. Homework assigned but not graded or commented on generates an effect size of only .28; however, the effect size increases to .78 when homework is graded, and to .83 (a percentile gain of 30 points), when the teacher provides written comments.

Readers who wish to locate more meta-analyses may wish to search the ERIC database for relevant citations by using the terms META ANALYSIS OR EFFECT SIZE in one set, and an intervention or instructional technique (e.g., COOPERATIVE LEARNING, COMPUTER ASSISTED INSTRUCTION) in another.

Some Caveats

The technique of meta-analysis injects useful scientific rigor into reviews of educational research, but it also has some limitations. For example, a synthesis of research studies may provide evidence about the overall effectiveness of a method, but not the specific details that would help guide implementation, such as the following raised by Marzano, Pickering, and Pollock (2001):

- Are some instructional strategies more effective in certain subject areas?
- Are some instructional strategies more effective at certain grade levels?
- Are some instructional strategies more effective with students from different backgrounds?
- Are some instructional strategies more effective with students of different aptitude? (p. 9)

A good meta-analysis may point to the need for more focused and refined research studies to answer those questions. It's also important to remember that it is provisional—representing the best evidence at the time it was conducted, but subject to change in the face of the evolution of the knowledge base.

References

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.

Cooper, H. (1989). Synthesis of research on homework. *Educational Leadership, 47* (3): 85-91.

Cooper, H., Lindsay, J.J., Nye, B., & Greathouse, S. (1998). Relationships among attitudes about homework, amount of homework assigned and completed, and student achievement. *Journal of Educational Psychology, 90* (1): 70-83.

Cooper, H., Valentine, J.C., Nye, B., & Lindsay, J.J. (1999). Relationship between five after-school activities and academic achievement. *Journal of Educational Psychology, 91* (2): 369-378.

Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Educational Research*, 5: 3-8.

Glass, G. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5: 351-379.

Marzano, R.J., Pickering, D.J., and Pollock, J.E. (2001). *Classroom Instruction That Works: Research-Based Strategies for Increasing Student Achievement.* Alexandria, VA: Association for Supervision and Curriculum Development.

National Institute of Child Health and Human Development (2000). Report of the National Reading Panel. Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction: Reports of the Subgroups (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office. [Online]. Available: http://www.nichd.nih.gov/publications/nrp/report.htm

Redfield, D.L., and Rousseau, E.W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research*, *51*(2): 237-245.

Walberg, H.J. (1999). Productive teaching. In H.C. Waxman and H.J. Walberg, eds. *New Directions for Teaching Practice and Research*, 75-104. Berkeley, CA: McCutchen Publishing Corporation.

This publication was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract ED99CO0032. The opinions expressed in this report do not necessarily reflect the positions or policies of OERI or the U.S. Department of Education.

Permission is granted to copy and distribute this ERIC/AE Digest.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[ ] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[X] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").